

YIHAO SUN

✉ yihao.sun@mila.quebec • www.yihaosun.cn • [Google Scholar](#)

EDUCATION

Ph.D. Student in Computer Science

May 2025 - Present

University of Montreal, Quebec, Canada
Mila Quebec AI Institute

M.Sc. in Computer Science and Technology

Sep 2021 - Jun 2024

Nanjing University, Jiangsu, China
LAMDA Group, School of Artificial Intelligence

B.Eng. in Software Engineering

Sep 2017 - Jun 2021

Sichuan University, Sichuan, China
School of Software Engineering

RESEARCH INTERESTS

My research interests center on reinforcement learning for general-purpose embodied intelligence, with a particular focus on world models and model-based reinforcement learning. Specifically, I am interested in building world models based on advanced generative models for embodied environments, and in developing MBRL algorithms that leverage such models for efficient policy learning, planning, and adaptation.

EXPERIENCE

Sharpa, Shanghai, China: Algorithm Engineer Intern

Sep 2024 – Nov 2024

- Achieving RL-based dexterous in-hand manipulation.
- Exploring a sample-efficient offline-to-online RL framework, and improving success rates from 40%+ to 90%+ with about 50 episodes online interaction.

REPRESENTATIVE PUBLICATIONS

Model-Bellman Inconsistency for Model-based Offline Reinforcement learning. ([pdf](#)|[code](#))

ICML'23

Co-first author, **Canonical citation on offline MBRL, 100+ citations.**

- Proposing a novel uncertainty quantification for uncertainty-driven model-based offline RL, which can better estimate the ideal uncertainty quantification, i.e., the Bellman error.
- Achieving SOTA on 20 of 27 evaluated benchmark datasets. Compared with MOPO, a representative model-based offline RL algorithm, the performance of our algorithm is improved by 42.4%.

Towards Practical World Model-based Reinforcement Learning for Vision-Language-Action Models. ([pdf](#)|[code](#)) ICML'26

Co-first author, **got OpenAI's interest!**

- Proposing a very clean yet efficient method for VLA-MBRL, including (i) adapting unified multimodal models (UMMs) for data-efficient world modeling; (ii) an interleaved view decoding mechanism to enforce multi-view consistency; and (iii) chunk-level branched rollout to mitigate error compounding.
- Showing better sample efficiency on 4 LIBERO simulation suites and 5 real-world tasks.

OPEN SOURCE PROJECTS

VLARLKit: An elegant PyTorch VLA-RL library. ([code](#))

300+ stars

100% code contribution.

- Simple and clear implementation — cleanly separated policy, rollout, runner, and model layers with minimal abstraction; easy to read, modify, and extend for research purposes.
- Though with simple implementation, it's still faster than other popular infras like RLinf (about 1.3x).
- Async off-policy training — supports asynchronous off-policy training, enabling non-blocking data collection alongside model updates.

OfflineRL-Kit: An elegant PyTorch offline reinforcement learning library. ([code](#))

380+ stars

95% code contribution.

- An elegant offline reinforcement learning library based on pure PyTorch.
- State-of-the-art offline RL algorithms, including model-free and model-based approaches.
- The first reproduction that matches or even exceeds the performance in the original papers of model-based offline RL.

AWARDS AND HONORS

Best Paper Award @ ICML Workshop on Decision-Making from Offline Datasets to Online Adaptation July 2026

Xiaomi Outstanding Scholarship (10 recipients schoolwide) Sep 2023

First Prize Scholarship in Nanjing University Nov 2021

First Prize Scholarship in Sichuan University Nov 2018

Excellent Student in Sichuan University Nov 2018

FULL PUBLICATION LIST

** denotes equal contribution.*

1. Zhilong Zhang*, Haoxiang Ren*, Yihao Sun*, Yifei Sheng, Haonan Wang, Haoxin Lin, Zhichao Wu, Pierre-Luc Bacon, and Yang Yu. Towards Practical World Model-based Reinforcement Learning for Vision-Language-Action Models. ICML'26.
2. Zhichao Wu*, Junyin Ye*, Zhilong Zhang*, Yihao Sun, Haoxin Lin, Jiaheng Luo, Haoxiang Ren, Lei Yuan, and Yang Yu. Speedup Patch: Learning a Plug-and-Play Policy to Accelerate Embodied Manipulation. ICML'26.
3. Lu Li, Tianwei Ni, Yihao Sun, and Pierre-Luc Bacon. The Three Regimes of Offline-to-Online Reinforcement Learning. ICML workshop, 2026 (Best paper).
4. Haoxin Lin, Siyuan Xiao, Yi-Chen Li, Zhilong Zhang, Yihao Sun, Chengxing Jia, and Yang Yu. ADM-v2: Pursuing Full-Horizon Roll-out in Dynamics Models for Offline Policy Learning and Evaluation. ICLR'26.
5. Zhilong Zhang et al. Hierarchical Value-Decomposed Offline Reinforcement Learning for Whole-Body Control. ICLR'26.
6. Yihao Sun, Zhilong Zhang, Yang Yu, and Pierre-Luc Bacon. Planning with Unified Multimodal Models. ICLR workshop, 2026.
7. Haoxin Lin, Yu-Yan Xu, Yihao Sun, Zhilong Zhang, Yi-Chen Li, Chengxing Jia, Junyin Ye, Jiaji Zhang, and Yang Yu. Any-step Dynamics Model Improves Future Predictions for Online and Offline Reinforcement Learning. ICLR'25.
8. Zhilong Zhang*, Tian Xu*, Xinghao Du*, Xingchen Cao, Yihao Sun, and Yang Yu. Improving Reward Model Generalization from Adversarial Process Enhanced Preferences. ICML'25.
9. Zhilong Zhang*, Ruifeng Chen*, Junyin Ye*, Yihao Sun, Pengyuan Wang, Jingcheng Pang, Kaiyuan Li, Tianshuo Liu, Haoxin Lin, Yang Yu, and Zhi-Hua Zhou. WHALE: Towards Generalizable and Scalable World Models for Embodied Decision-making. NeurIPS workshop, 2025.
10. Haoxin Lin, Hongqiu Wu, Jiaji Zhang, Yihao Sun, Junyin Ye, and Yang Yu. Episodic Return Decomposition by Difference of Implicitly Assigned Sub-trajectory Reward. AAAI'24.
11. Zhilong Zhang*, Yihao Sun*, Junyin Ye, Tianshuo Liu, Jiaji Zhang, and Yang Yu. Flow to Better: Offline Preference-based Reinforcement Learning via Preferred Trajectory Generation. ICLR'24.
12. Ruifeng Chen*, Xiong-Hui Chen*, Yihao Sun, Siyuan Xiao, Minhui Li, and Yang Yu. Policy-conditioned Environment Models are More Generalizable. ICML'24.
13. Tian Xu*, Zhilong Zhang*, Ruishuo Chen, Yihao Sun, and Yang Yu. Provably and Practically Efficient Adversarial Imitation Learning with General Function Approximation. NeurIPS'24.
14. Yihao Sun*, Jiaji Zhang*, Chengxing Jia, Haoxin Lin, Junyin Ye, and Yang Yu. Model-Bellman Inconsistency for Model-based Offline Reinforcement Learning. ICML'23.
15. Haoxin Lin*, Yihao Sun*, Jiaji Zhang, and Yang Yu. Model-based Reinforcement Learning with Multi-step Plan Value Estimation. ECAI'23.